Program Evaluation (PPHA-34600, Winter 2020)
Problem Set 2
Due: Wednesday, February 05, 2020 at 8 p.m.

Instructions:

This problem set consists of three files: (1) This document with instructions and questions, (2) an article, "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size" (Angrist and Evans 1998), and (3) a dataset similar to that used in the Angrist and Evans (1998) analysis. Use the article and data provided to answer the questions below.

All problem set submissions must be made through Canvas before the deadline. No late submissions will be accepted. In your submission you should provide a write up where you answer the questions below, selectively cutting and pasting output where needed. Be concise in your write up; excess wordiness will be penalized. You should also submit a log file that includes commands and results for your entire analysis. (This means that you will upload 2 files: one ending in .pdf, and one text code file.) You may use either Stata or R. Remember that your grade mostly comes from the written report—your code or log is a "replication" file that shows you did the work. Imagine that the grader will not look at your code until they are done reading the write up, and that is how you should type it up. If you put it all together in a R notebook, knit it to .pdf before submission.

Questions:

1. (10 pts) Suppose you were interested in the broad question, *What is the effect of the number of children on how much a woman works?* Suppose "one more child" is the treatment, use words and the potential outcomes framework to describe the ideal comparison for answering this question. Be sure to be clear about the unit of analysis. (Hint: Your answer should contain the phase "one more child", potential outcomes, and the ideal measure of an impact.)

2. (15 pts) Use words and describe:

   (a) what the before-after estimator would compare for this question.

   (b) what the naive estimator would compare for this question.

   (c) two kinds of bias that are a problem given these estimate.

3. (25 pts) Consider the comparison that Angrist and Evans make using the instrument of having two same sex children, and the dependent variable hours worked per week among all women.

   (a) What is the estimated treatment parameter? To whom does it apply?

   (b) Describe for the context who are the compilers, defiers, always-takers and never-takers.

(c) Describe for the context what does the exclusion restriction requires. How is this different if the dependent variable is married women's husbands' income?

(d) Describe for the context what the monotonicity assumption means.

4. (5 pts) Suppose that a policymaker wants to use this research to assess whether the government should provide a subsidy to help families with 3 or more children. Discuss what this research can say about such a policy. What are the limitations?

Data Cleaning:

5. (15 pts) We will analyze the data set where a woman is the unit of analysis, but the data set has an entry for each child. (report your code for these tasks in your writeup)

   (a) Prepare your dataset for analysis where you have 1 observation per mother. (Hint this is sometimes referred to a reshaping a long dataset into a wide dataset.)

   (b) Generate four indicators: one for when a mother has more than 2 kids (2pluskids), on for when a mother has two children of the same sex (samesex), one for when the mother has twins at her second birth (twins2nd), and one for when the mother has less than 12 years of education (nodiploma)

Data Analysis:
The following questions ask you to estimate and analyze the effect of having more than 2 children on the number of hours a woman works per week.

6. (10 pts) Using OLS with heteroskedasticity robust standard errors, estimate the relationship between having more than two children on the average number of hours that a mother works per week, using just your indicator variable. Report results and comment.

7. (10 pts) Now add controls for the (relative) age of the mother (assume the data if from 1980), if she has less than 12 years of education, and if she has a college degree. Report results. What changes? Comment on the relationship between hours worked and education, be sure to include the omitted category. Can you interpret this causally?

8. (15 pts) Now suppose that having two children of the same sex is a good predictor of having more than 2 children. Use the samesex variable as an instrument for having more than 2 kids.

(a) Report the first stage results when you include the controls from above. Comment on (all) the necessary assumptions for samesex to be a valid instrument to estimate the effect of a woman having more than two children on hours worked per week.

(b) Report the second stage results. What is the treatment parameter recovered. Who is the effect for? What do you find the effect is? Can you interpret this causally?

9. (15 pts) Now use having twins at the second birth as an instrument for having more than 2 kids.

   (a) Include the controls discussed earlier and write the first and second stage equations you will estimate.

   (b) Report the first stage results. Comment on the strength of the instrument.

   (c) Report the second stage results. What is the treatment parameter recovered. Who is the effect for? What do you find the effect is? How does this differ from the estimate recovered using samesex as the instrument. Discuss possible reasons these effects could be different.

References

Angrist, Joshua D., and William N. Evans. 1998. "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size." *The American Economic Review* 88, no. 3 (January): 450–477.